

# RooStats for Searches

Grégory Schott, *on behalf of the RooStats team*

KIT, Institut für Experimentelle Kernphysik, Karlsruhe, Germany

## Abstract

The RooStats toolkit, which is distributed with the ROOT software package, provides a large collection of software tools that implement statistical methods commonly used by the High Energy Physics community. The toolkit is based on RooFit, a high-level data analysis modeling package that implements various methods of statistical data analysis. RooStats enforces a clear mapping of statistical concepts to C++ classes and methods and emphasizes the ability to easily combine analyses within and across experiments. We present an overview of the RooStats toolkit, describe some of the methods used for hypothesis testing and estimation of confidence intervals and finally discuss some of the latest developments.

## 1 Introduction

The RooStats project [1, 2] is a collaborative open source project initiated by members of ATLAS, CMS and the CERN ROOT team. The RooStats toolkit — based on previously existing code used in ATLAS [3] and CMS [4], which has been extended and improved — has been distributed with ROOT since summer 2008. The toolkit provides and consolidates statistical tools needed for LHC analyses and allows one to apply and compare the most popular and well-established statistical approaches. Thanks to readily available well-known tools, results across experiments can be better understood and compared. This is not only a desirable feature but also a required one when it comes to combining analysis results as will be discussed later. Finally, the RooStats project aims to provide reasonably flexible, well-tested, documented tools. The RooStats developments benefit from scientific oversight from the statistics committees of both experiments.

In High Energy Physics, the goal of an analysis is usually to test a prediction or search for new physics, leading to the estimation of the statistical significance of a possible observation or the construction of confidence intervals — often expressed as upper or lower limits in case of a non-observation. The most common statistical procedures are:

- point estimation: i.e., the determination of the best estimate of parameters of the model,
- confidence or credible interval estimation: i.e., regions representing the range of parameters of interest compatible with the data,
- hypothesis tests: i.e., comparing the data to two or more hypotheses,
- goodness of fit: to quantify how well a given model describes the observed data.

RooStats aims to cover some of these common statistical procedures.

The RooStats package is built on top of RooFit [5], which is a data modeling toolkit developed originally within the BaBar collaboration and now integrated into ROOT. The most crucial element of RooFit is its ability to model probability densities, likelihood functions, and data, in a very flexible way that can deal with arbitrarily complex cases. Some recent developments in RooFit provide additional tools specifically needed by RooStats. The RooStats code is organized into three groups of classes: *calculators* that perform the statistical calculations, *results* and *utilities* that facilitate the RooStats work flow.

After a few generalities, given in Sect. 2, the classes implementing statistical inferences and results are discussed in Sect. 3. In Sect. 4, we describe RooStats utilities, while Sect. 5 will have a few words on some applications and perspectives.

## 2 Generalities

We begin by clarifying some of the terminology commonly used:

- *Observables*: quantities that are measured by an experiment (e.g., mass, helicity angle, output of a neural network) that form a *data set*.
- *Model*: the probability density function (PDF) — either parametric or non-parametric — that describes one or multiple observables and normalized so that their integral over any observable is unity.
- *Parameters of interest*: parameters of the model whose value we wish to estimate or constrain (e.g., a particle mass or a cross-section).
- *Nuisance parameters*: uncertain parameters of the model other than the ones of interest (e.g., parameters associated with systematics, such as normalization or shape parameters). The treatment of nuisance parameters varies according to the statistical approach.

### 2.1 Likelihood Function

The modeling of the likelihood function is the principal task of RooFit. RooFit, which builds on ROOT, maps mathematical concepts to RooFit classes. For example, variables, functions, probability densities, integrals, a space point, or a list thereof, are handled by RooRealVar, RooAbsReal, RooAbsPdf, RooRealIntegral, RooArgSet and RooAbsData, respectively. A large collection of functions are available to describe the PDF. The functions are handled by classes inheriting from RooAbsPdf and can be easily combined to build arbitrarily complex models through addition, multiplication, and convolution. For both data and models there exist some binned and unbinned representations. For each model, integration and maximum likelihood fitting is supported and utilities are provided for the Monte Carlo generation of pseudo data, in order to perform "toy" studies, and for the visual inspection of results. The utilities and great modularity of RooFit are the principal factors that drove the choice of RooFit as the basis of RooStats. One can work with arbitrarily complex data and models and one can handle large sets of observables and parameters.

Most statistical methods usually start with a likelihood function. A rather general likelihood function, for use in our field, with multiple observables, can be written as:

$$L(\mathbf{x}|r, s, b, \theta_s, \theta_b) = e^{-(rs+b)} \prod_{j=1}^n [rs f_s(\mathbf{x}_j|\theta_s) + b f_b(\mathbf{x}_j|\theta_b)]. \quad (1)$$

The PDFs  $f_s$  and  $f_b$  represent the distributions of observables  $\mathbf{x}$  for the signal and background, with parameters  $\theta_s$  and  $\theta_b$ , respectively. The parameters  $s$  and  $b$  — typically, the expected signal and background counts, respectively — are constrained by the number  $n$  of observed events<sup>1</sup>. In this likelihood function a strength factor  $r$  multiplies the expected number of signal events<sup>2</sup>.

### 2.2 Model Configuration

Before one can perform a statistical inference, it is necessary to specify the model: the PDF of possible observables, the actual observables, the parameters of interest, the nuisance parameters, possibly a Bayesian prior, etc. The RooStats calculators can be configured, via the constructor, either with the model specifications given as individual RooFit objects or with a ModelConfig object, in which the

<sup>1</sup>Sometimes described as an extended likelihood; it can also be viewed as the limit of a binned multi-Poisson likelihood function with arbitrarily small bins.

<sup>2</sup>This is sometimes done to redefine the parameter of interest such that  $r$  is the ratio of the signal production cross-section to the expected value of the cross-section. For example, in the search for the Standard Model Higgs boson, obtaining a 95% CL upper-limit for  $r = 1$  means the Standard Model Higgs hypothesis can be excluded at 95% CL.

model specification is bundled. For most of the calculators both configuration mechanisms are available. The idea behind `ModelConfig` is to provide a uniform way to configure calculators. The downside is that it becomes less obvious what elements of the `ModelConfig` are necessary for a given calculator. For example, the prior probability will not be used in frequentist-based calculations while the list of observables, which is mainly used to generate pseudo-data, is not needed when computing Bayesian limits.

The model is often completed by a set of observed data. Moreover, the calculators can be configured for a number of options specific to the statistical algorithms (e.g., number of Monte Carlo iterations, size of the test, test statistic, etc.). Finally, the calculator is run and returns the result of a hypothesis test or a confidence interval.

### 3 RooStats Calculators

Below, we describe the RooStats calculators, which are based on the following conceptual approaches:

- *Classical or Frequentist*: this school of statistics restricts itself to statements of the form "probability of the data given the hypothesis". Probability is interpreted as a limit of relative frequencies of various outcomes.
- *Bayesian*: this school of statistics views probability more broadly, which permits statements of the form "probability of the hypothesis given the data". Typically, probability is interpreted as a "degree of belief" in the veracity of an hypothesis.
- *Likelihood*: this approach uses a frequentist notion of probability (e.g., it does not require the specification of a prior for the hypothesis), but inferences are not guaranteed to satisfy some frequentist properties (e.g., coverage). Like the Bayesian approach, this likelihood approach obeys the likelihood principle, while frequentist methods do not.

We give a brief description of the methods available in RooStats and refer the reader to textbook literature for details (see, for example [6, 7]).

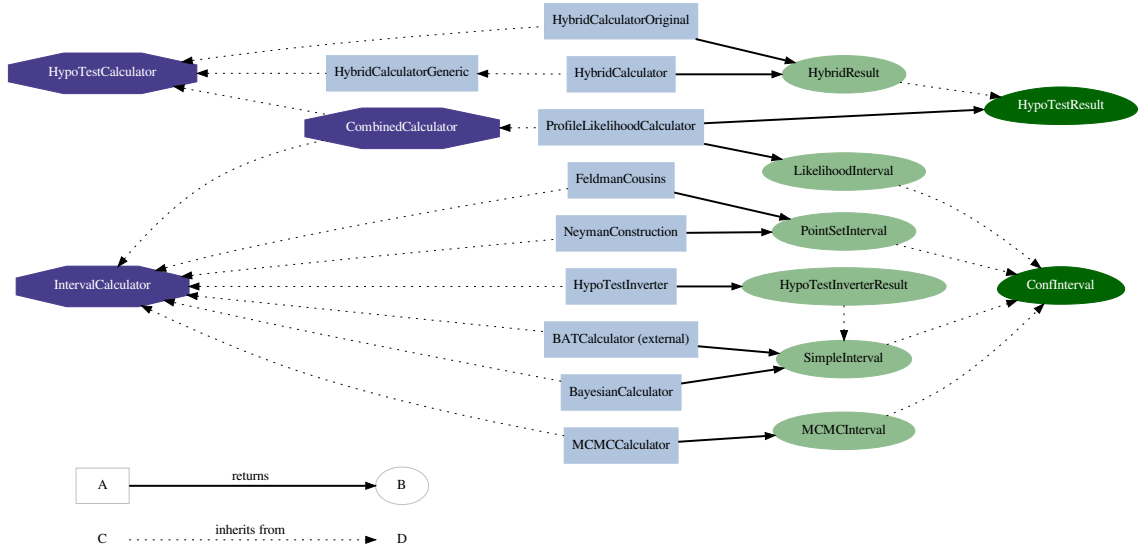
As can be seen from Fig. 1, there are two general classes of calculators in RooStats: those performing hypothesis-tests and those computing confidence or credible intervals, which inherit, respectively, from the classes `HypoTestCalculator` and `IntervalCalculator` and return, respectively, objects inheriting from the classes `HypoTestResult` or `ConfInterval`.

The `IntervalCalculator` interface allows the user to provide the model, the data set, the parameters of interest, the nuisance parameters and the size  $\alpha$  of the test ( $\alpha = 1 - \text{CL}$ , where CL is the confidence/credible level). After configuring the calculator, a `ConfInterval` pointer is returned via the method `IntervalCalculator::GetInterval()`. Depending on the calculator used, a different type of `ConfInterval` will be returned (e.g., connected interval, multi-dimensional interval, etc.) but each shares the ability to test if a point lies within the interval using the method `ConfInterval::IsInInterval(p)`.

The `HypoTestCalculator` can be configured with the model, the data and parameter sets specifying the two hypotheses to be tested. Through `HypoTestCalculator::GetHypoTest()`, a pointer to the result can be retrieved and the result object can be queried for  $p$ -values and the corresponding significances, or  $Z$ -values, found by equating a  $p$ -value to a one-sided Gaussian tail probability and solving for the number of standard deviations. In this convention, a  $p$ -value of  $2.87 \times 10^{-7}$  corresponds to a  $Z$ -value of  $5\sigma$ .

#### 3.1 Profile-Likelihood Calculator

The `ProfileLikelihoodCalculator` class implements a likelihood-based method to estimate a confidence interval and to perform an hypothesis test for a given parameter value. To illustrate the method, let



**Fig. 1:** Diagram of the interfaces for hypothesis testing and confidence interval calculations and classes used to return the results of these statistical tests.

us assume that the likelihood function depends on a set  $K$  parameters  $\theta$ , one of which is the parameter of interest. From the likelihood function  $L(\mathbf{x}|\theta_0, \theta_{i \neq 0})$ , similar to the one of Eq. (1) but where the parameter of interest  $r$  has been renamed  $\theta_0$ , for generality, the profile likelihood function is the numerator in the ratio:

$$\lambda(\theta_0) = \frac{L(\theta_0, \hat{\theta}_{i \neq 0})}{L(\hat{\theta}_0, \hat{\theta}_{i \neq 0})}. \quad (2)$$

The denominator,  $L(\hat{\theta})$  is the absolute maximum of the likelihood, while the numerator is the maximum value of the likelihood for a given value of  $\theta_0$ .

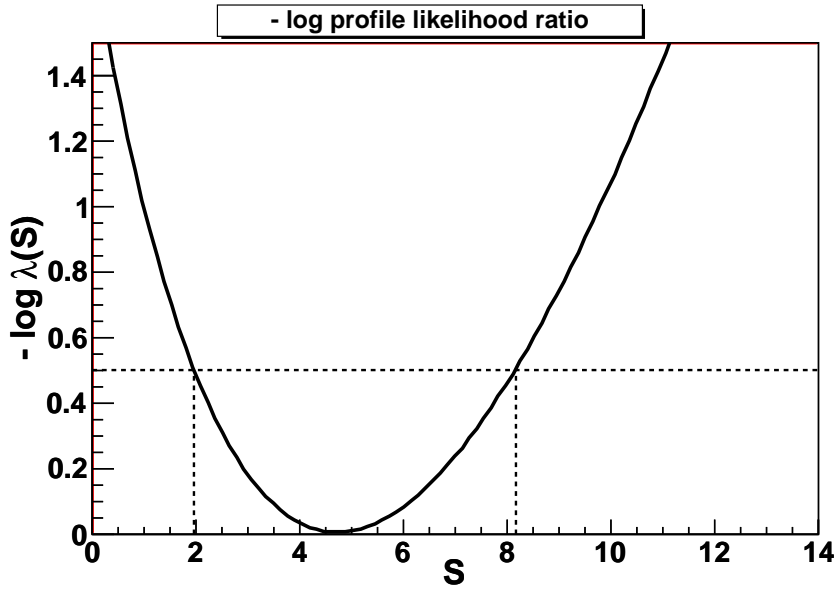
Under certain regularity conditions, Wilks's theorem demonstrates that asymptotically  $-2 \ln \lambda(\theta_0)$  follows a  $\chi^2$  distribution. In the asymptotic limit, the likelihood ratio test statistic  $\lambda(\theta_0)$  has a parabolic shape:

$$-2 \ln \lambda(\theta_0) = -2(\ln L(\theta_0) - \ln L(\hat{\theta}_0)) = n_\sigma^2, \text{ with } n_\sigma = \frac{\theta_0 - \hat{\theta}_0}{\sigma}, \quad (3)$$

where  $n_\sigma$  represents the number of Gaussian standard deviations associated with the parameter  $\theta_0$ . From this construction, it is possible to obtain the one- or two-sided confidence intervals (see Fig. 2). Owing to the invariance property of the likelihood ratios, it can be shown that this approach remains valid for non parabolic log-likelihood functions. This method is also known as MINOS in the physics community, since it is implemented by the MINOS algorithm of the Minuit program. Given the fact that asymptotically  $-2 \ln \lambda$  is distributed as a  $\chi^2$  variate, an hypothesis test can also be performed to distinguish between two hypotheses characterized by different values of  $\theta_0$ .

In this approach, systematic uncertainties are taken into account by augmenting the likelihood function with terms that encode the knowledge we have of the systematic uncertainties and the profiling is now done over all nuisance parameters including those for the systematics.

This likelihood-based technique for estimating an interval and performing a hypothesis test is provided in RooStats by the `ProfileLikelihoodCalculator` class. The class implements both the `IntervalCalculator` and `HypoTestCalculator` interfaces. When estimating an interval, this calculator returns a `LikelihoodInterval` object, which, in the case of multiple parameters of interest, rep-



**Fig. 2:** Plot of the log profile likelihood curve as function of the parameter of interest,  $\theta_0 \equiv S$ . The  $1\sigma$  interval (68% CL) is obtained from the intersect of the  $-\log \lambda$  curve with the horizontal dashed line  $-\log \lambda = 0.5$ .

resents a multi-dimensional contour. When performing a hypothesis test, a `HypoTestResult` object is returned with the significance for the null hypothesis. Another class exists, `LikelihoodIntervalPlot`, to visualize the likelihood interval in the case of one or two parameters of interest (as shown in Fig. 2). A newly developed class, `ProfileInspector`, allows inspection of the value of the nuisance parameters for each value of the parameter of interest along the profile log-likelihood curve.

### 3.2 Bayesian Calculators

Bayes theorem relates the probability (density) of a hypothesis given data to the probability (density) of data given a hypothesis. The inversion of the probability is achieved by multiplying the likelihood function (the probability of the data given an hypothesis) by a prior probability for the model, which is characterized by parameters of interest and, typically, one or more nuisance parameters. This product is normalized so that the integral of the posterior density, over all parameters, is unity. The calculation of credible intervals, that is, Bayesian confidence intervals, requires the calculation of the cumulative posterior distribution. In the Bayesian approach, nuisance parameters are removed by marginalization, that is, by integrating over their possible values. RooStats provide two different types of Bayesian calculator, the `BayesianCalculator` and `MCMCCalculator` classes, depending on the method used for performing the required integrations.

The current implementation of the `BayesianCalculator` class works for a single parameter of interest and uses numerical integration to compute the posterior probability distribution. Various algorithms provided by ROOT for numerical integration can be used, including those based on Monte Carlo integration, such as implemented in the programs Vegas or Miser. The result of the class is a one-dimensional interval (`SimpleInterval`) obtained from the cumulative posterior distribution.

The `MCMCCalculator` uses a Markov-Chain Monte Carlo (MCMC) method to perform the integration. The calculator runs the Metropolis-Hastings algorithm, which can be configured by specifying parameters such as the number of iterations and burn-in-steps, to construct the Markov Chain. Moreover, it is possible to replace the default uniform proposal function with any other proposal function. The result of the `MCMCCalculator` is a `MCMCInterval`, which can compute the confidence interval for the desired

parameter of interest from the Markov Chain. The `MCMCInterval` integrates the posterior density from its mode downwards until the interval has a  $1 - \alpha$  probability content<sup>3</sup>. The `MCMCIntervalPlot` class can be used to visualize the interval and the Markov chain.

Users can also input the RooStats model into the Bayesian Analysis Toolkit (BAT) [8], a software package that implements Bayesian methods via Markov-Chain Monte Carlo. In the latest release, BAT provides a class, `BATCalculator`, which can be used with a similar interface to the RooStats `MCMCCalculator` class. Developments are foreseen that will further integrate BAT within RooStats.

### 3.3 Neyman Construction

The Neyman construction is a pure frequentist method to construct an interval at a given confidence level,  $1 - \alpha$ , such that coverage is guaranteed for fully-specified probability models. A detailed description of the method is given in Ref. [6]. RooStats provides a class, `NeymanConstruction` that implements the construction. The class derives from `IntervalCalculator` and returns a `PointSetInterval`, a concrete implementation of `ConfInterval`.

The Neyman construction requires the specification of an ordering rule that defines the order in which potential observations are to be added to the interval in the space of observations until the desired confidence level is reached. The ordering rule is usually specified in terms of a specific test statistic. Consequently, the RooStats class must be configured with this information before it can produce an interval. More information can now be provided with the introduction of the interfaces `TestStatistic`, `TestStatSampler`, and `SamplingDistribution`. Different test statistics are available, including:

- Simple likelihood ratio:  $Q = L_1(\theta_0 = 1)/L_0(\theta_0 = 0)$ ,
- Ratio of profiled likelihoods:  $Q' = L_1(\theta_0 = 1, \hat{\theta}_{i \neq 0})/L_0(\theta_0 = 0, \hat{\theta}'_{i \neq 0})$ ,
- Profile likelihood ratio:  $\lambda(\theta_0) = L_1(\theta_0, \hat{\theta}_{i \neq 0})/L_0(\hat{\theta}_0, \hat{\theta}_{i \neq 0})$ .

Another aspect to decide is how to sample it: assuming asymptotic distribution, generating toy-MC experiments with nuisance parameters fixed (used in `NeymanConstruction`) or with nuisance parameters sampled according to a prior distribution (used in `HybridCalculator`).

Common configurations, such as the Feldman-Cousins approach — where the ordering is based on the profile likelihood ratio as the test statistic [9], can be enforced by using the `FeldmanCousins` class. A generalization of the Feldman-Cousins procedure, when nuisance parameters are present, generating toy Monte Carlo experiments with nuisance parameters fixed as described in [3, 10], is also available.

The Neyman construction considers every point in the parameter space independently. Consequently, there is no requirement that the interval be connected nor that it have a particular structure. The result consists of a set of scanned points labeled according to whether they are inside or outside the interval (`PointSetInterval` class). The user either specifies points in the parameter space that are to be used to perform the construction or a range and a number of points within the range, which will be scanned uniformly in a grid. For each scanned point, the calculator will give the sampling distribution of the chosen test statistic. This is typically obtained by toy Monte Carlo sampling, but other techniques exist and can, in principle, be used. In particular, newly developed code may be helpful when testing hypotheses with very small  $p$ -values through the application of importance sampling techniques.

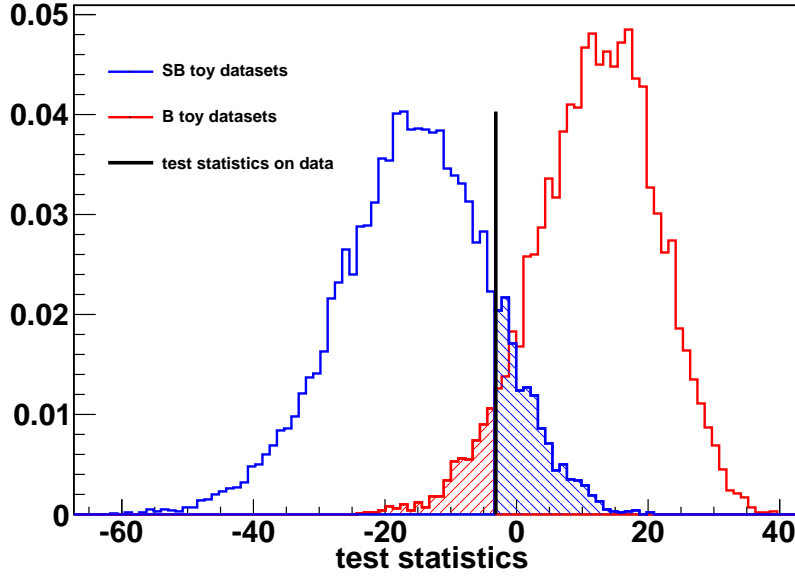
### 3.4 Hybrid Calculator

This calculator implements a Bayesian/frequentist hybrid approach for hypothesis testing. It consists of a frequentist toy Monte Carlo method, as in the Neyman construction, but with a Bayesian marginalization of nuisance parameters [11]. This technique is often referred to as a "Bayesian-Frequentist Hybrid".

---

<sup>3</sup>It should be noted that these *highest posterior density intervals* are not invariant under one-to-one reparametrisation.

For example, let us define the null hypothesis,  $H_0$ , to be the background-only or no signal hypothesis, and  $H_1$  to be the alternate hypothesis that a signal is present along with background. In order to quantify the degree to which each hypothesis is favoured or excluded by the experimental observation, one chooses a test statistic which ranks the possible experimental outcomes. Given the observed value of the test statistic, the  $p$ -values,  $CL_{sb} \equiv p_1$  and  $CL_b \equiv 1 - p_0$ , can be computed. Since the functional forms of the test statistic distributions are typically not known a priori, a large number of toy Monte Carlo experiments are performed in order to approximate these distributions. Figure 3 provides an example of such distributions from the two pseudo data sets and where the observed value of the test statistic lies.



**Fig. 3:** Result from the hybrid calculator, the distributions of a test statistic in the background-only (red, on the right) and signal+background (blue, on the left) hypotheses. The vertical black line represents the value obtained on the tested data set. The shaded areas represent  $1 - CL_b$  (red) and  $CL_{sb}$  (blue).

Systematics uncertainties are taken into account through Bayesian marginalization. For each toy Monte Carlo experiment, the values of the nuisance parameters are sampled from their prior distributions before generating the toy sample. The net effect is to broaden the distribution of the test statistic, as expected in the presence of systematic uncertainties, and thus degrade the separation of the hypotheses.

This procedure is implemented in RooStats by the `HybridCalculator` class. The input to the class are the models for the two hypotheses, the data set and, optionally, the prior distribution for the nuisance parameters, which is sampled during the toy generation process. As for the `NeymanConstruction`, the test statistic can be freely parameterized. The results of the `HybridCalculator` consists of the test statistic distribution for the two hypothesis, from which the hypothesis  $p$ -value and associated  $Z$ -value can be obtained. Since the simulation of the distributions could be computationally expensive, RooStats permits different results to be merged, which makes it possible to run the calculator in a distributed computing environment. The `HybridPlot` class provides a way of plotting the result, as shown for example in Fig. 3.

By varying the parameter of interest representing the hypothesis being tested (for example, the signal cross-section) one can obtain a one-sided confidence interval (e.g., an exclusion limit). RooStats provides a class, `HypoTestInverter`, which implements the interface `IntervalCalculator` and performs the scanning of the hypothesis test results of the `HybridCalculator` for various values of one parameter of interest. By finding where the confidence level curve of the result intersects the de-

sired confidence level, an upper limit can be derived, assuming the interval is connected. An estimate of the computational uncertainty is also provided. Finally, when defining exclusion limits, the condition that defines the upper bound can be chosen: either one can use the  $p$ -value  $p_1$  of the alternate hypothesis (the pure-frequentist approach) or the ratio of  $p$ -values  $CL_s = p_1/(1 - p_0)$  (modified-frequentist approach [12]).

## 4 RooFit and RooStats Utilities

### 4.1 RooFit's Workspace

One element of RooFit whose addition has been driven by the development of the RooStats project (although it would still be useful even without RooStats) is the `RooWorkspace` class. It is a container for RooFit objects that can be written to a ROOT file. When a RooFit object is imported from a file (e.g., a complex PDF with multiple parameters), all the other dependent objects are imported too. Later, it is very easy to rebuild and initialize all the parameters, to reconstitute the original PDF, via a single recall from the `RooWorkspace` (while still permitting adjustments to the imported object). These features make it possible to save the complete likelihood function, as well as the data, to a file in a well defined fashion, either as a technical convenience, as an intermediate step towards the combination of the results of multiple analyses or for the grander purpose of electronic publication of these results. In addition, the `RooWorkspace` interfaces to a newly developed utility, `RooFactoryWSTool`, which permits the building of a large class of RooFit objects in an interpreted mode with an intuitive syntax based on strings. Multiple dependent parameters are also defined, created and stored in the `RooWorkspace` on-the-fly, thereby allowing, for example, the creation of a Gaussian PDF in one line, instead of the four needed to create one (the PDF along with its observable and two parameters) using the RooFit classes directly. It will be discussed later how this factory tool is complemented by RooStats' `HLFactory` class.

### 4.2 User-Friendly Model Specification

Tools that simplify and automate the description of complex models in a user-friendly way are usually referred to as model factories. There are currently two such utilities provided within RooStats: `HLFactory` and `HistFactory`. Their use is optional. For more experienced users or in more complex cases, direct use of the lower level RooFit classes may be preferred.

`HLFactory` is a RooStats class whose aim is to disentangle the C++ code doing the calculations from the physics-driven and analysis-specific description of the probability models. The later can be written to a single text file describing all (and only) the physics inputs that are to be processed later in a single line of code. The fact that `HLFactory` is built as a simple wrapper around the `RooWorkspace` factory utility sidesteps the need to define yet another language that a user would have to learn, while not restricting the application to specific analyses since this model factory supports everything the `RooWorkspace` factory does. In addition, python-like instructions are added that allow better structuring of the description (through includes) and along with comments on the analysis model. Finally (and optionally), the `HLFactory` also allows the easy combination of multiple channels to form a combined model and combined data set.

`HistFactory` is a collection of classes to handle template histogram-based or binned analyses. It allows such analyses to use RooStats without requiring knowledge of the RooFit modeling language; instead, the likelihood function and elements of the statistical analysis are specified through an XML configuration file, which is used to produce the model. In this approach, the user provides histogram templates of one observable and of models for different contributing samples (e.g., of the signal and background processes). Then, the normalization in terms of number of events for each of these channels can be decomposed — for example, as a product of luminosity, efficiency, cross-section terms — each of which can be affected by systematic uncertainties. It supports Gaussian, gamma and log-normal distributions for nuisance parameters. Finally, histograms of variations can be provided that specify the



related systematic changes. Multiple channels can be given and combined and parameters which are identical across channels can be easily identified.

### 4.3 Other Utilities

Not all utilities are listed in this document. Here we mention briefly three more:

- `SPlot`, a class implementing a technique used to produce weighted plots of an observable distribution in a multi-dimensional likelihood-based analysis [13].
- `RooNonCentralChiSquare`, a class in `RooFit` that outlines the use of a generalization of Wilks' theorem called Wald's theorem which states that the asymptotic distribution of the test statistic  $\lambda(\mu)$  for  $\mu \neq \mu_{true}$  is a non-central  $\chi^2$  [14],
- `BernsteinCorrection`, a class that augments the nominal probability with a positive-defined polynomial given in the Bernstein basis, which can be used as an approach to incorporate systematic effects in a PDF.

## 5 Statistical Combinations and Perspective

The combination of results is a commonly used method for improving sensitivities or measurements of signals. With `RooStats`, the combination can be performed at the analysis level in contrast to combinations performed at the level of published results. This means that the global likelihood function for the ensemble of the analyses to be combined is explicitly written and the statistical analysis is performed on this combined likelihood. This approach has advantages, such as being able to account for known correlations consistently. But, it also has its inconvenience, such as making the likelihood function a quite complex object. One strong motivation for the `RooStats` project was to simplify the process of combining analyses by providing a tool that allows this to be done simply for arbitrarily complex models.

In December 2010, ATLAS and CMS created the LHC-HCG group mandated to prepare and produce a combined Higgs result from the LHC (with similar efforts also on-going in other analysis groups within the collaborations). `RooStats` will be used for the combination and one of the first tasks of the group has been to complement its validations with comparison to results obtained from independent software in specific analysis cases<sup>4</sup>. While the validations appear satisfactory so far, the `RooStats` team will keep improving interfaces and fix performance issues as well as develop new complementary tools based on users' experiences and feedback.

One aspect of statistical data analysis is left open by `RooStats`, namely that of the choice of statistical method. In that respect, it allows the implementation of one recommendation of the ATLAS and CMS statistics committees, which is that various methods be applied and compared (although different methods are not expected to give the same results since they have different properties and provide answers to different questions). A more specific method and statistical procedure to use when combining ATLAS and CMS analyses is a topic still under discussion and one of the focuses of this PHYSTAT conference.

### Acknowledgements

The `RooStats` contributors are thankful to the members of the ATLAS and CMS statistics committees for the exchange of ideas, advice and encouragement. I also wish to thank L. Lyons and the rest of PHYSTAT committee for the organization of the very rich and useful conference and the invitation to present there progress on the development of the `RooStats` toolkit.

---

<sup>4</sup>For further insights on these activities see Ref. [15]

## References

- [1] L. Moneta *et al.*, *The RooStats project*, PoS ACAT2010, 057 (2010) [arXiv:1009.1003].
- [2] RooStats homepage: <https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome>.
- [3] K. S. Cranmer, *Statistics for the LHC: Progress, challenges and future*, proceedings of PHYSTAT 2007, CERN-2008-001, 47 (2007).
- [4] D. Piparo, G. Schott and G. Quast, *RooStatsCms: a tool for analysis modelling, combination and statistical studies*, J. Phys. Conf. Ser. 219, 032034 (2010) [arXiv:0905.4623].
- [5] W. Verkerke, *Statistical software for the LHC*, proceedings of PHYSTAT 2007, CERN-2008-001, 169 (2007).
- [6] F. James, *Statistical methods in experimental physics, 2nd edition*, Word Scientific (2006).
- [7] K. Nakamura *et al.* *The Review of Particle Physics - Chapter 33*, J. Phys. G37, 075021 (2010).
- [8] A. Caldwell, D. Kollar, K. Kröninger, *BAT: The Bayesian Analysis Toolkit*, Comput. Phys. Commun. 180, 2197 (2009) [arXiv:0808.2552].
- [9] G. Feldman and R. D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Phys. Rev. D57, 3873 (1998).
- [10] K. S. Cranmer, *Frequentist hypothesis testing with background uncertainty*, proceedings of PHYSTAT 2003 [physics:0310108].
- [11] R. D. Cousins and V. L. Highland, *Incorporating systematic uncertainties into an upper limit*, Nucl. Instrum. Meth. A320, 331 (1992).
- [12] A. L. Read, *Modified frequentist analysis of search results (The CLs method)*, CERN OPEN-2000-205 (2000).
- [13] M. Pivk and F. R. Le Diberder, *SPlot: A statistical tool to unfold data distributions*, Nucl. Inst. Meth. A555, 356 (2005) [physics:0402083].
- [14] G. Cowan *et al.*, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C71, 1554 (2011) [arXiv:1007.1727].
- [15] K. S. Cranmer, *Combining ATLAS and CMS Higgs searches*, proceedings of PHYSTAT 2011 (these proceedings).